

University of Groningen

BiNet

Dhali, Maruf A.; Wit, Jan Willem de; Schomaker, Lambert

Published in:
ArXiv

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Early version, also known as pre-print

Publication date:
2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Dhali, M. A., Wit, J. W. D., & Schomaker, L. (2019). BiNet: Degraded-Manuscript Binarization in Diverse Document Textures and Layouts using Deep Encoder-Decoder Networks. *ArXiv*.
<https://arxiv.org/pdf/1911.07930v1>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

BiNet: Degraded-Manuscript Binarization in Diverse Document Textures and Layouts using Deep Encoder-Decoder Networks

A PREPRINT

Maruf A. Dhali*, Jan Willem de Wit, Lambert Schomaker

Department of Artificial Intelligence, Bernoulli Institute
University of Groningen
The Netherlands

November 20, 2019

ABSTRACT

Handwritten document-image binarization is a semantic segmentation process to differentiate ink pixels from background pixels. It is one of the essential steps towards character recognition, writer identification, and script-style evolution analysis. The binarization task itself is challenging due to the vast diversity of writing styles, inks, and paper materials. It is even more difficult for historical manuscripts due to the aging and degradation of the documents over time. One of such manuscripts is the Dead Sea Scrolls (DSS) image collection, which poses extreme challenges for the existing binarization techniques. This article proposes a new binarization technique for the DSS images using the deep encoder-decoder networks. Although the artificial neural network proposed here is primarily designed to binarize the DSS images, it can be trained on different manuscript collections as well. Additionally, the use of transfer learning makes the network already utilizable for a wide range of handwritten documents, making it a unique multi-purpose tool for binarization. Qualitative results and several quantitative comparisons using both historical manuscripts and datasets from handwritten document image binarization competition (H-DIBCO and DIBCO) exhibit the robustness and the effectiveness of the system. The best performing network architecture proposed here is a variant of the U-Net encoder-decoders.

Keywords Document binarization · Historical manuscripts · Pattern recognition · Artificial intelligence · Computer vision · Deep learning · Convolutional neural network · Conditional generative adversarial network · Dead Sea Scrolls

1 Introduction

In a digitized image of a handwritten document, the ink-based pixels are the result of a physical ink deposition process, where a surface material absorbs the pigment creating the foreground. In contrast, the original unaffected material texture appears as the background. In a typical handwritten document,

$$N_{ink} \ll N_{background}$$

where, N_{ink} is the number of ink pixels and $N_{background}$ is the number of background pixels.

The binarization process of handwritten document-image allocates a binary value to each pixel of the image [1]; 0 for ink and 1 for not-ink. Thus, this process separates the foreground (meaningful information, in general, the ink) from the background (the surface material). The binarized images are compressed and pose significant importance towards analyzing the document. It facilitates character recognition, segmentation, and transcription pipeline [2, 3, 4]. Further processing of the handwriting towards writer identification and script-style development also depends on the success of the binarization process itself [5, 6]. Over the years, many techniques have been proposed to perform binarization tasks. However, it is always challenging to obtain good results due to the diversity of the handwritten

*Corresponding author;

Address: Nijenborgh 9, 9747 AG Groningen, The Netherlands; Email: m.a.dhali@rug.nl, web: www.rug.nl/staff/m.a.dhali/

documents. One technique may perform well for some specific type of documents but fail for other types. The problem becomes challenging when it comes to historical manuscript binarization [7].

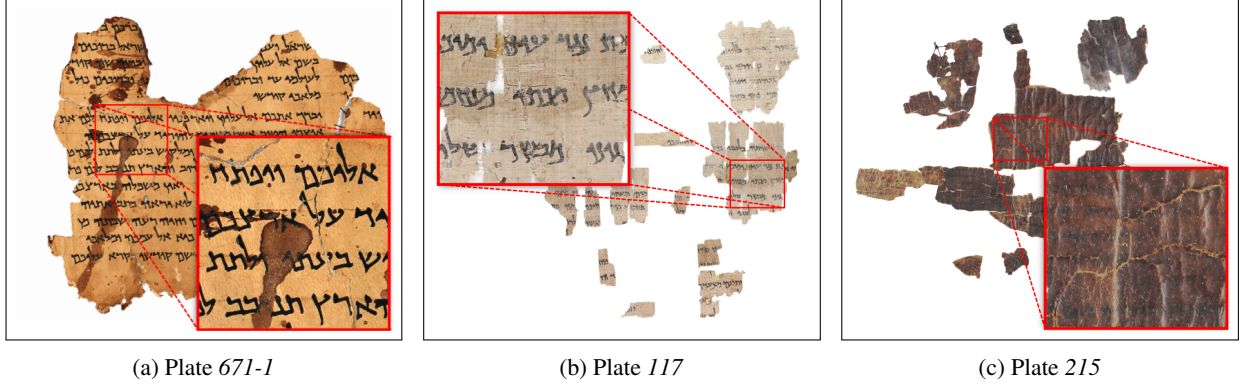


Figure 1: Three images from the Dead Sea Scrolls collection show the diversity of the materials and their current state with difficult readability due to various degradation. They are all RGB-colored images of the plates (physical arrangements of the fragmented materials on a plane surface). Plate 607-1 contains only one physical fragment, whereas the next two plates (117 and 215) contain multiple physical fragments. Fragments from plate 117 were produced from papyrus, and the repetitive patterns of the fibers are visible in the zoomed-in section of the image. The other two plates contain fragments made from parchments.

Numerous historical manuscripts collections are residing all over the world [8]. Most of them are significant, both culturally and scientifically [9]. The Dead Sea Scrolls (DSS) are one such collection. They are ancient manuscripts discovered in the mid-20th century in the Judean Desert, in between Jerusalem and the Dead Sea. Most were written over a period of almost four centuries (ca. 250 BC to ca. 135 AD) and hold tremendous historical, religious, and linguistic significance [10]. The recent digitization of this collection has opened the door for pattern recognition techniques to be applied to revise existing hypotheses on the writers and dates of these scrolls [5]. However, these documents have diverse document textures and are heavily degraded (see Figure 1), mostly due to the materials, natural aging, the preservation processes, and the places they were kept in. In order to perform pattern recognition techniques on the original content (texts) only, the images need to be preprocessed. One of the critical steps in this preprocessing is a binarization technique with the ability to keep the original content of these documents as intact as possible.

There are several challenges in binarizing the DSS images. Similar to many other historical manuscripts, the DSS collection profoundly suffers from document degradation problems. Due to aging and natural causes, the individual glyphs (characters) of the DSS images often show fading effects. Some of the images also show thinning of the characters along with broken (missing or completely faded) parts. Some images often suffer from uneven illumination problems due to the surface material [11]. On top of all these, the most severe issue is the low contrast between ink and background (see Figure 2).

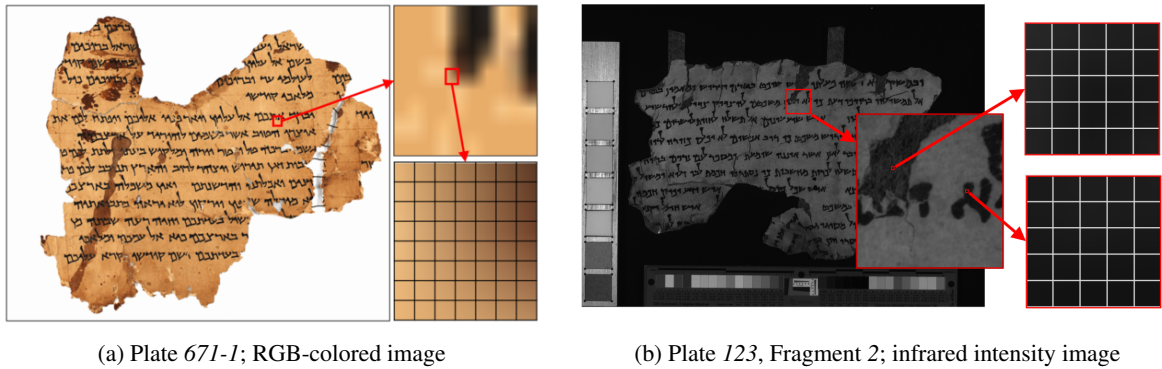


Figure 2: The degraded DSS documents show the variation of contrast. The *left* image-set shows an RGB-colored plate with zoomed in to pixel level at the edge of an ink-stroke. The *right* image-set shows an IR-image (captured in 924nm wavelength of light) of a fragment in grayscale. Here, the two zoomed-in portions from the degraded background (top) and the ink (bottom) show no visual differences in color intensity.

1.1 Why binarization is still important

Although many modern deep-learning methods in document analysis can be trained end-to-end, directly with a grayscale or color image, this is not desirable in an e-science approach for the humanities studies. For instance, a direct end-to-end solution on clustering the DSS collection can be achieved for writer identification and script-style evolution, but there is always a risk of getting the solution for the wrong cause as in the ‘Russian (hidden) tank problem’ [12]. The decision of a neural network may be based on spurious correlations with the texture of the background of different materials, such as papyrus and parchment. The fiber statistics of papyrus manufacturing batches may add to the wrong cause, as well. Additionally, irrelevant materials like the background, rice papers, number tags, scale bars, color calibrators, and other patterns must not be allowed to contribute to the process. So binarization is necessary, and it needs to be precise.

Many of the existing binarization techniques are pixel-intensity based. It is already clear that any method working on the pixel-intensity only will struggle in producing excellent results for the DSS images. An intelligent binarization method should accommodate all different variability in the DSS collection and still provide excellent results. Semi-automatic selection of the region of interest or a manual one-off preprocessing technique may obtain good binarization, but it will not serve as a robust solution for the whole corpus. For objective analysis, including writer identification and dating, a non-biasing foreground-background separation is required. Separating foreground-background can be a severe problem to solve but significant for scholarly analysis. This problem can be illustrated using a suggestion put forward by the eminent palaeographer Ada Yardeni. She ascribed fifty-seven or possibly even ninety-three manuscripts to one scribe ([13, 14]). Two of such fragment images are shown in Figure 3. This figure also presents the binarization results from three of the most famous traditional techniques. In the case of ink separation, the techniques perform considerably well in some regions of the image but fail in most of the remaining areas. Now, for the case of writer identification, the palaeographer who is strictly interested in comparing two hand-writings to check the hypothesis from Yardeni would not accept any of the results based on such binarization data. For writer identification, the binarization technique should be capable enough to focus on the original written content only, not the surrounding material, not even the markings and scale bars.

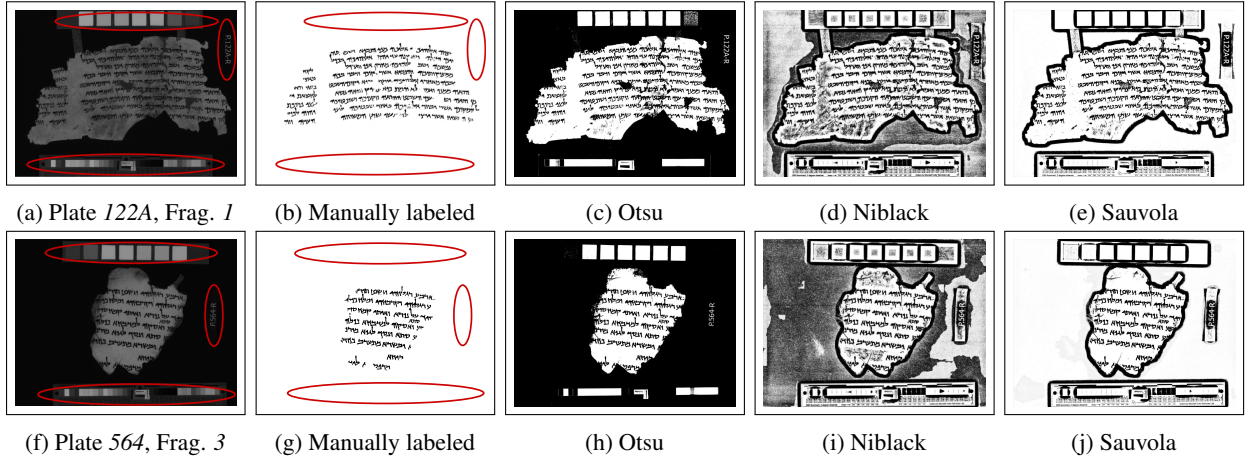


Figure 3: An illustration of popular binarization techniques applied directly to two of the DSS fragment-images. Sub-figures (a) and (f) show the original IR-images (captured in 924-nm wavelength of light). Sub-figures (b) and (g) show the corresponding manually labeled ground truths by human experts. The red-circled areas show the parts where the human experts ignore the irrelevant contents of the images, such as the color-calibration bars, scales, and numbers. The binarization results of techniques proposed by Otsu [15], Niblack [16], and Sauvola [17] are presented for both the fragment images. All these three methods fail to provide output images that focus only on the original contents, unlike a human expert.

1.2 Goals

Intensity alone is not a sufficient heuristic in the binarization task. Rather than using a single filtering technique, a bank of trainable filters is needed to solve this problem. The system should be able to ignore the irrelevant information during binarization and should be able to include everything which is part of the original content. A system needs to be adequately intelligent to focus on writing like a human does with reasonable accuracy (see Sub figures 3b and 3g). An artificial neural network, with suitable architecture requiring a small amount of training data, can be the right solution as a multi-filter trainable method for these diverse materials. Towards the goal of both robustness and obtaining

optimal results, this article proposes BiNet, an unbiased automatic end-to-end binarization approach for handwritten documents based on the U-Net architecture ([18]). Inspired by the works of ‘pix2pix’, a general-purpose solution for image-to-image translation problems using conditional adversarial networks ([19]), the proposed model includes the encoder-decoder structure without the discriminator part making it a variant of the U-Net. Skip-connections are added between the contracting and the expansive path by simply concatenating all channels from one layer to the other. This concatenation circumvents the bottle-neck issue at the deepest layers of the encoder and ensures the precise positioning of the foreground-background pixels. A simple illustration of the proposed network is provided in Figure 5.

This study demonstrates the effectiveness of the proposed model on the binarization task using the collection of the DSS images. Both the RGB-colored images and the pseudo-colored images are used. Pseudo-colored images are fused from grayscale intensity-images of different spectral bands. A simple technique is proposed in generating ground-truth images to train the network. Similar to many historical manuscripts, the DSS collection also lacks in ground-truth labels. It is time consuming and tedious to create ground truth at the pixel level. So the work in this paper ensures that the training data is precise, includes a sufficient amount of variability, and the network can perform well with a small amount of training data. On top of this, transfer learning techniques are introduced so that the proposed model becomes usable for different collections. Additionally, the Handwritten Document Image Binarization Competition (H-DIBCO) datasets are tested to exhibit the effectiveness of the system. Quantitative and qualitative results are presented to compare it with other techniques. BiNet, the proposed model, shows better performance with robustness to the variability of the data. Overall, this article makes the following contributions:

- BiNet: the complete framework for a fully automated binarization method for DSS images that allows further analysis of the original contents of the collection.
- A network capable of learning the differentiation between relevant and irrelevant information during the training process, and thus providing intelligent and useful binarized outputs.
- An in-depth analysis of the proposed binarization tool using comparative studies and quantitative analysis.
- Multi-purpose usability of the binarization tool for different manuscript collections (including H-DIBCO images) using effective transfer learning techniques.
- A simple and easy technique to generate precise ground-truths (training images) for the DSS collection that can be extended to any degraded historical manuscripts.
- A new way to generate fused (pseudo-color) images from multi-spectral bands to yield more information (qualitatively) than any of the individual bands.

2 Related Works

Document image binarization is one of the most common research problems that has been addressed numerous times in the field of document analysis. Some of these methods have achieved great success in many applications and have become popular over time. Otsu [15] is one of the most commonly used methods. This unsupervised and non-parametric method automatically selects a global threshold based on the grayscale histogram of a given image with no prior information. The Otsu method is one of the simplest binarization methods that perform well when the image is qualitatively clean with a uniform background. Unfortunately, most of the historical manuscripts do not contain a uniform background or a clear bimodal pattern. Thus, a global thresholding approach is not suitable for these types of documents [20]. A gradual change in the uniformity of the background can be handled using small local patches of the target image by local adaptive thresholding. Several local thresholding methods have been developed, such as Niblack [16], Sauvola [17], local max-min by Su et al. [21], and AdOtsu [22]. Descriptive statistics (mean and standard deviation) are calculated on the local area of a pixel to obtain the local threshold. This local thresholding technique is then performed over the whole target image. It performs well compared to the global thresholding techniques, but often shows poor performance in the case of historical manuscripts where the documents are highly degraded with extremely non-uniform backgrounds. To extend further, in cases similar to the DSS collections, both the global and local thresholding methods fail to provide useful results (an example of this can be found in Figure 3 from the previous section).

In order to improve the results of threshold-based binarization, several image-processing techniques are used as an enhancement part of the document-processing pipeline along with the binarization itself. Shi et al. used the mathematical morphological operator and region-growing techniques [23]. In order to compute the final threshold, a Wiener filter [24] is used for the background surface by Gatos et al. [25]. Instead of using the Wiener filtering, robust regression is used by Vo et al. for document binarization of noisy and non-uniform background [26]. Phase-derived features are used for ancient document image binarization in the works of Nafchi et al. [27]. In order to enhance and reconstruct degraded documents, a method using non-local patch means (NLPM) is proposed by Moghaddam et al. [28]. Bio-inspired models

have already been used for text detection in natural images [29]. Similarly, models based on the OFF-center ganglion cells of the human visual system is used for the improvement of document enhancement and binarization [30]. Different contrast enhancements are performed to adjust local grayscale contrast to improve the binarization results compared to traditional threshold-based techniques on DIBCO datasets [31].

Many previous works on binarization have exploited the prior knowledge of texts in the document. The edge pixels of the texts can be extracted by techniques similar to the Canny edge detector [32]. This technique is already proposed by Chen et al. in their double threshold image binarization method [33]. One generalization of edge pixels is transition pixels with extreme transition values. These pixels are calculated on a small neighborhood using the intensity difference, and then the gray-intensity threshold is calculated from the statistical information of the transition set [34]. On the contrary, structural symmetric pixels (SSPs) are used to determine local thresholds in a neighborhood, and a voting system is utilized for multiple thresholds for the binarization task [35]. Automatic parameter tuning can be done by utilizing a global energy function inspired by a Markov random field model by incorporating edge discontinuities [36]. All these methods are mostly based on traditional image processing and pattern recognition techniques and may have promising characteristics. However, they are designed to attain good results on certain types of documents and lack in addressing the diversified degradation problems similar to the DSS collection with its wide spectrum of writing-surface materials.

With the success of deep learning in sophisticated image understanding [37], several neural network architectures have been proposed for handwritten document binarization and analysis. A fully convolutional network (FCN) is proposed, which operates at multiple image scales starting from full-resolution [38]. A convolutional encoder-decoder is used by Peng et al. [39] on the LRDE document binarization dataset [40, 41]. In the recent works of Calvo-Zaragoza et al., a selectional autoencoder is used for the binarization task [1] on a couple of DIBCO datasets [42, 43], Balinese palm leaf manuscripts [44], Persian documents from PHI [45], and music notations from SAM and ES [46]. Afzal et al. have proposed the use of the recurrent neural network for document binarization [47]. This work has been extended by Westphal et al. [48] by using Grid Long Short-Term Memory (Grid LSTM) networks [49] for the binarization task. A hierarchical deep supervised network (DSN) architecture is proposed, which shows better performance than the Grid LSTM on DIBCO datasets [50]. In the recent work of He et al., an iterative deep learning technique is proposed for document enhancement and binarization [51] on several DIBCO datasets, the Bickley-diary dataset [52], the PHIDB dataset [53], and the SynchroMedia Multi-spectral dataset [54].

All these neural network-based techniques are useful, and they present improved performances in many cases. However, most of the time, the datasets used do not pose extreme cases of degradation along with diverse material textures like parchment and papyrus (see Figure 4). A texture modeling can be performed using the Markov random field (MRF) [55], but it will be extremely complicated in the case of DSS images. Explicit foreground and background modeling have been proposed by Sriman et al. in the classification of text blocks in the scene images [56], but the binarization task needs precise localization of each of the ink pixels. On top of this, in the case of the fragment images of the DSS collection, the unnecessary elements and modern number tags should be ignored in the outcome of the binarization (as discussed in the previous section; see Figure 3). The desired binarization tool, on the one hand, should be robust enough to handle extremely degraded historical manuscripts written in parchments and papyrus, like the DSS collection, and, on the other hand, should perform well in general document cases including the DIBCO images.

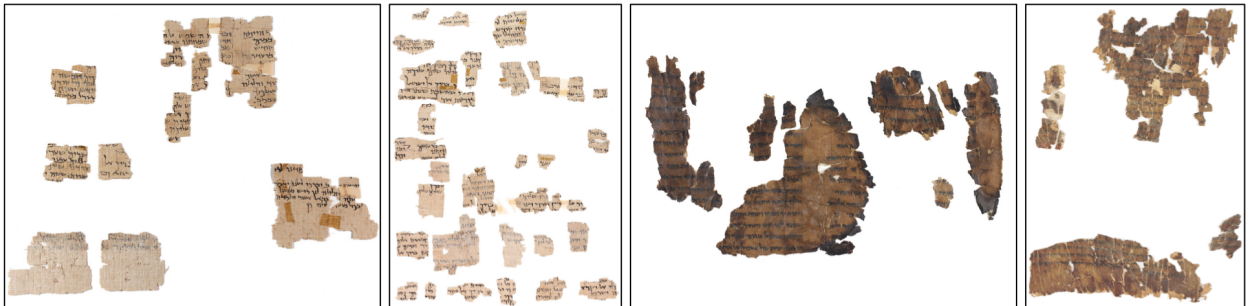


Figure 4: The evidence for ink versus background resides only partly in the intensity of an individual pixel. The evidence is heavily present in external features (of papyrus and parchment), as well. The figure shows four RGB-color images of full plates from the DSS collection (from *left to right*: 463A, 464, 1080, and 1082). The first two have papyrus as a surface material for writing, and the latter two have parchment. Both the materials show distinctive degradation and decaying of characters (inks). The binarization tool should be able to find the separation of ink from the surface materials explicitly.

Although many of the previous works already provided us with different tools for benchmark datasets, a robust tool is yet to be designed that performs not only outstanding binarization for severe cases like the DSS collection but also shows consistent performance in general cases. The implementations of U-Net [18] and pix2pix [19] methods are particularly relevant here. Though U-Net was initially designed for biomedical-image segmentation, this has been used already for accurate pixel classification in one of the recent competitions on document image binarization (DIBCO 2017 [57]). On the other hand, pix2pix is proposed initially as a general-purpose solution to image-to-image translation problems using conditional adversarial networks [58, 59] and is not designed to perform document binarization tasks. However, the inspiration lies in the performance of the pix2pix method on image-to-image translation tasks with highly-structured graphical outputs by learning a loss adapted to the task. The proposed method in this article is based on a similar idea where the outputs are precise and straightforward representations of highly complex inputs. Hence, the proposed model is inspired by pix2pix and is a variant of the general U-Net approach.

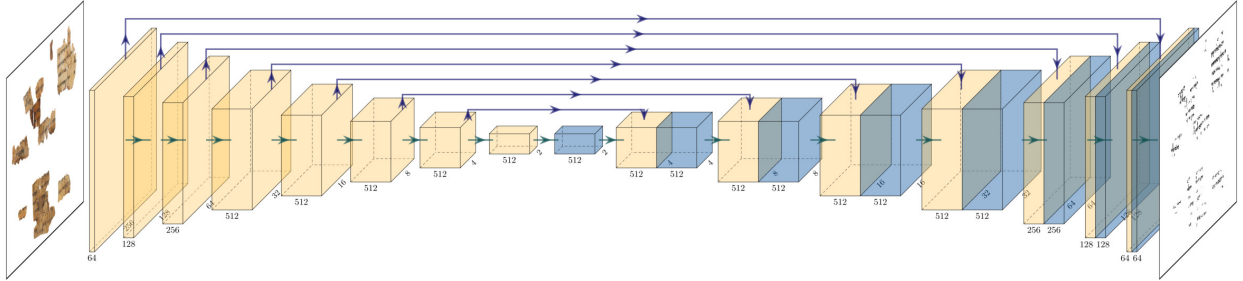


Figure 5: The proposed network architecture shows the encoder (contracting path) at the left-half and the decoder (expanding path) at the right-half of the image. Each step in the decoder part receives a concatenation with the corresponding feature-map from the encoder part through the skip connections. This concatenation circumvents the bottle-neck issue at the deepest layers of the encoder and ensures the precise localization of the foreground-background pixels.

3 Methodology

In this section, we will briefly explain the proposed model and the complete methodology. We will present the BiNet framework, the network architecture with technical details, hyperparameters, the transfer learning techniques, and the datasets.

3.1 BiNet

The document image binarization task is simply a two-class classification problem. For handwritten documents, we define the two classes as foreground and background. The simple solution is to decide labels for each of the pixels of an image, one by one. The background study already shows several methods working in this direction. However, for the cases of extremely degraded historical manuscripts, this is a problematic direction, even with local thresholding in a small neighborhood. We need a trainable network that learns the target content and classifies each pixel by taking into account its neighbors in a local region. In this article, we propose a new approach utilizing a network architecture, the BiNet, that works end-to-end, providing the binarization in just one single step taking into account the knowledge of the original content. Rather than classifying each pixel separately, the BiNet efficiently works on the whole input image to generate a binarized output image of the same size.

The implementation of BiNet is a variant of the U-Net architecture [18] that is capable of doing complex binarization tasks. The original U-Net was designed for biomedical image segmentation with precise localization and can be trained end-to-end with very few images. Due to these two traits, we follow the typical shape of U-Net with skip connections to build our model. Thus, our model becomes an image-generator (which shares similarity with pix2pix [19] but differs in the adversarial parts) that produces a binarized image from an input DSS image. The original content (the texts; ground-truth) of a DSS image is accompanied by several factors including the texture of writing-surface, various degradation, and irrelevant materials (numbers, scale-bars, and the surface of the platform). The BiNet learns a mapping from the input image x to output image y :

$$x = y + \delta \tag{1}$$

where x is the DSS image as it is with degradation and other factors, y is the latent original content (ink from the original writing; the ground truth), and δ is the noise that comprises all the information except the original content. The

network is trained by minimizing the classification error from the $L1$ -loss function:

$$L_1 = \sum_{i=1}^n |y_{true} - y_{predicted}| \quad (2)$$

where n is the number of pixels in the input image, y_{true} is the ground-truth, and $y_{predicted}$ is the prediction from the network. As the binarized output is less complex than the input image, simply using the $L1$ -regression is enough. Additionally, the $L1$ loss function is less affected by the outliers, making it a preferable choice over $L2$ loss function for the DSS collection. The $L2$ loss has the tendency of producing the border effects.

Table 1: Detailed description of the BiNet architecture. Here, $Conv(f, h, w)$ denotes as a convolutional operator with f filters and $h \times w$ kernel sizes; $BNorm()$ refers to batch normalization; $Dropout(r)$ denotes dropout operation with ratio of r connections at each time; $LeakyReLU$ refers to leaky rectified linear unit; $Tanh$ refers to Tanh activation.

Original input	Input at Encoder	Encoding layers	Decoding layers	Final output
[0, 255] ^{256×256} or, [0, 255] ^{256×256×3}	256x256	Conv (64,4,4,2) Actv (LeakyReLU 0.2)	Conv (512,4,4,2) BNorm () Actv (LeakyReLU 0.2) Dropout (0.5)	[0, 1] ^{256×256}
	128x128	Conv (128,4,4,2) BNorm () Actv (LeakyReLU 0.2)	Conv (1024,4,4,2) BNorm () Actv (LeakyReLU 0.2) Dropout (0.5)	
	64x64	Conv (256,4,4,2) BNorm () Actv (LeakyReLU 0.2)	Conv (1024,4,4,2) BNorm () Actv (LeakyReLU 0.2) Dropout (0.5)	
	32x32	Conv (512,4,4,2) BNorm () Actv (LeakyReLU 0.2)	Conv (1024,4,4,2) BNorm () Actv (LeakyReLU 0.2)	
	16x16	Conv (512,4,4,2) BNorm () Actv (LeakyReLU 0.2)	Conv (1024,4,4,2) BNorm () Actv (LeakyReLU 0.2)	
	8x8	Conv (512,4,4,2) BNorm () Actv (LeakyReLU 0.2)	Conv (512,4,4,2) BNorm () Actv (LeakyReLU 0.2)	
	4x4	Conv (512,4,4,2) BNorm () Actv (LeakyReLU 0.2)	Conv (256,4,4,2) BNorm () Actv (LeakyReLU 0.2)	
	2x2	Conv (512,4,4,2) BNorm () Actv (LeakyReLU 0.2)	Conv (128,4,4,2) BNorm () Actv (LeakyReLU 0.2)	
			Conv (1,4,4) Actv (Tanh)	

3.2 Network Architecture

The network architecture is illustrated in Figure 5. The technical details of each of the layers can be found in Table 1. In the encoder part, we have Convolution-BatchNorm-LeakyReLU layers with different number of filters. The decoder part consists of Convolution-BatchNorm-Dropout-LeakyReLU with a 50% dropout in the first three layers, then the remaining layers consist of a Convolution-BatchNorm-LeakyReLU structure. We used the leaky rectified linear unit (LeakyReLU, [60]) as the activation function. The convolutions are 4×4 spatial filters applied with stride 2 and padding 1. The hyperparameters are set empirically through grid search. No max-pooling layer is used. Experimental results show that BiNet is one of the optimal topologies for DSS image binarization. The model works with a fixed image size of 256×256 and accepts both color (3 channel) and grayscale (1 channel) images, and always outputs a 256×256 binary image. Please note that the document image can be much larger and of variable sizes. The implementation processes all the input images by dividing them into equal pieces of 256×256 patches (padding is performed if necessary) to provide the input images to the network and combine the individual outputs to get the full binarized image.

Instead of ReLU, we used LeakyReLU to avoid collapsing gradient. Parametric ReLU has the same advantage with one difference that the slope of the output for negative inputs is a learnable parameter while in the Leaky ReLU it is a hyperparameter. Convolutional layers replace max-pooling with increased stride making it a sub-sampling step. The network is large enough for the DSS images to be trained on, and it can learn all the necessary invariances without using max-pooling layers, without any loss in accuracy [61].

3.3 Transfer Learning

The BiNet structure is initially proposed for binarizing the DSS fragment images. Now, in the case of full plate images, the whole model that is pre-trained from scratch using DIBCO images can be retrained to update the weights of the network. This retraining can be done using a small number of ground-truth plate images due to the high similarity of the plate images to the DIBCO images. A BiNet architecture can also be used in different historical manuscript collections using this simple transfer learning technique with only a small amount of training data.

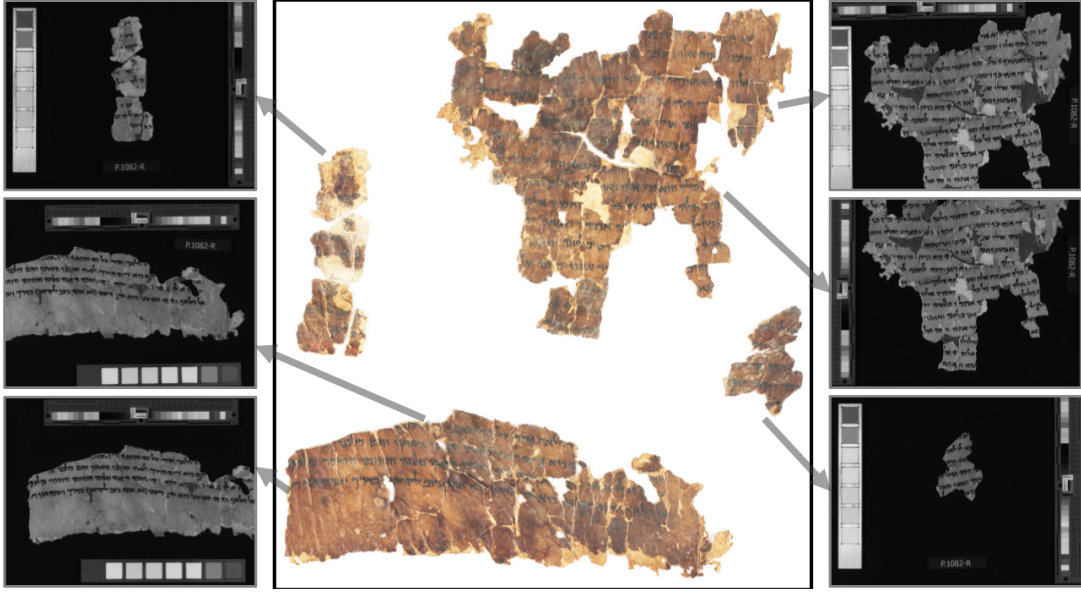


Figure 6: The RGB-color image of the full plate 1082 and the grayscale intensity images (924nm wavelength) of all the four fragments of this plate. Two separate images have been captured for each of the two large fragments to cover the entire areas.

3.4 Datasets

As our primary dataset, we will use the DSS image collection. There are several sources for the digital images of the DSS manuscripts. In this study, we will use the high-resolution multi-spectral images kindly provided to us by the Israel Antiquities Authority (IAA), which derive from their Leon Levy Dead Sea Scrolls Digital Library project. These images are offered to scholars and the general public on their website [62]. For each of the scrolls fragments, the IAA produces one color-image and several multi-spectral images on both recto and verso in 28 different exposures with a resolution of 1215 pixels per inch (PPI) at 1:1 ratio [63]. In addition to the fragment images, there are also color images of the full plates where the fragments are physically preserved (see Figure 6). Depending on the arrangement, a full plate may contain one fragment or several different fragments. In this study, we will first use the fragment images to train and test the model. Later, we will use the plate images through the transfer learning technique (as described in 3.3).

The **fragment images** have a dimension of 5412×7216 or 7216×5412 pixels, depending on the orientation of the physical fragment. We scaled this down to 50%, to speed up the training and testing process. The resulting dimensions are therefore 2706×3608 or 3608×2706 pixels respectively. The proposed network, BiNet, takes input image with a dimension of 256×256 . We, therefore, divide the input image into small-images of 256×256 pixels. This way, we end up with 165 small images per original image. As the dimensions of the images are not divisible by 256, the cuts from the edges of the images have a smaller size. We change these images to size 256×256 by padding them. This same procedure can be followed to accommodate any image size at the input of the network. We will train three models with different inputs of fragment images, so we can see which of the three images gives the best binarization result:

- RGB-color images (captured in visible light; $445nm$ wavelength)
- Grayscale intensity images (captured in infrared; $924nm$ wavelength)
- Fused images (details in Subsection 3.5)

The **plate images** have variable dimensions of $\sim 3000 \times 4000$ pixels. They all are RGB-color images. We follow the same procedure as in the case of fragment images to produce small-images of 256×256 from the plate images.

In addition to the DSS images, we will also use the publicly available **(H-)DIBCO** datasets from the document image binarization competitions of years 2009 through 2018 [64, 65, 66, 67, 42, 43, 57, 68]). Finally, in order to check the robustness of the system, we will use grid images produced from several different historical manuscripts (non-DSS) from the Monk system [69].

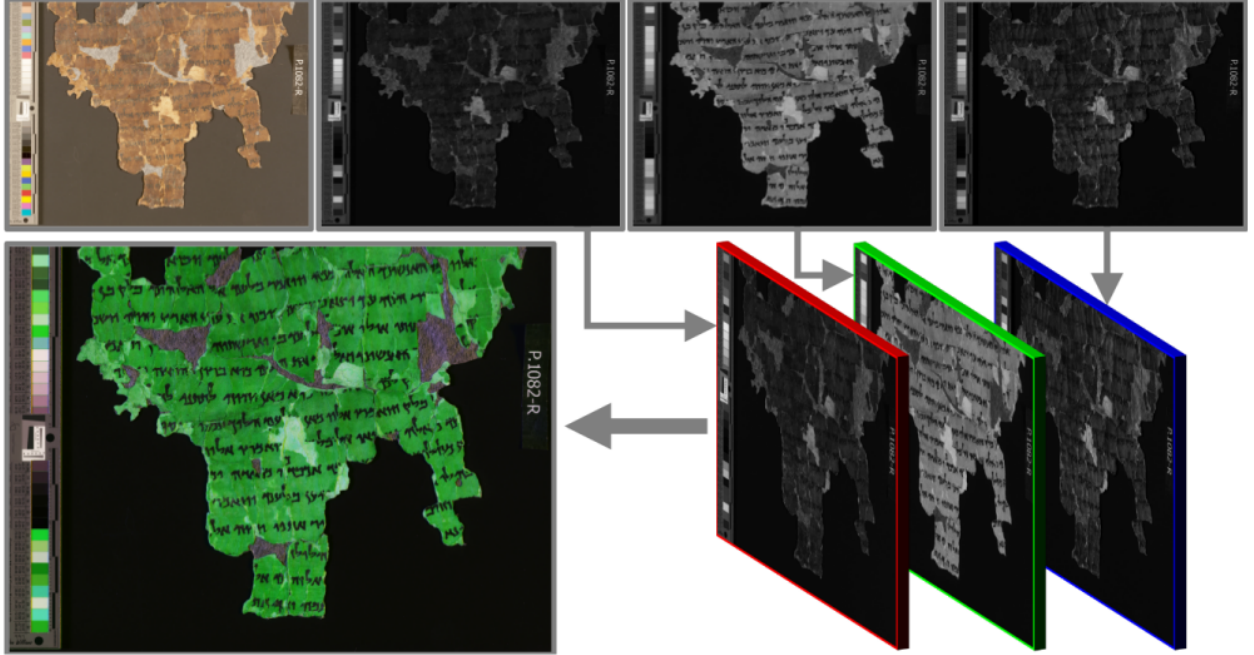


Figure 7: An illustration of the image fusion technique. In the top row: the first image from right is the RGB-color image of fragment 1 from plate 1082. The next three are the band images with wavelength of $595nm$, $924nm$, and $638nm$ respectively. In the bottom row: at the right side the formation of three channels from the grayscale images are shown, and at the left is the resultant pseudo-color image (fused image).

3.5 Image Fusion

We use the grayscale image resulting from the intensity of light at each pixel in the wavelength of $924nm$. This image is an interesting choice from our end. We select this particular wavelength because the resulting grayscale image shows the maximum visible contrast between the ink and the background compared to any other wavelengths. To extend our work further and to improve the ink-background separation, we take advantage of other multi-spectral band-images. We propose here an image-fusion technique to create a pseudo-color image. We take grayscale intensity images from three separate wavelengths: $595nm$, $638nm$, and $924nm$. With these three images, we produce a new fused image (pseudo-color image) with three channels. Image with wavelength $595nm$ goes to the R-channel, $924nm$ to the G-channel, and $638nm$ to the B-channel. Figure 7 shows an example of three grayscale images and the resulting pseudo-color RGB-image from image-fusion. By doing this, we hope to capture more details that emerge from various lighting conditions to improve the binarization result.

3.6 Ground Truth

One of the biggest challenges in working with DSS images is the lack of ground-truths. Besides, the DSS collection is not a structured or complete dataset, like many other historical manuscripts. In order to use trainable networks, we first need labeled train images. To create the ground-truths, we used GIMP, a free and open-source raster graphics editor

(version 2.8.16 [70]). To establish the credibility of the ground-truths, palaeographic experts labeled the images (see Section 7).

We have proposed a simple method for the labeling task using the GIMP tool. First, a transparent layer is created on the top of the fragment-image that would be used for training the network. This layer is of the same dimension as the original image. Now, by zooming into the vicinity of the characters, the palaeographic experts mark the inks-pixels in red, capturing the characters from the bottom layer (original image) to the transparent layer (new image: the ground truth) with a pen of 1 pixel accuracy. This work is similar to creating the carbon copy of a document, but the other way around. In this task, the palaeographic expert will overwrite the entire original content on the transparent layer. Due to choosing a pen size of 1 pixel, we ensure that only the ink pixels are marked red, and everything else remains transparent. Once the task is complete, the transparent layer is taken out and saved as a separate image where the red pixels are converted to black and everything else to white (ground truth). As the task is done using a computer mouse, it is simple but time-consuming. It should be noted that the palaeographic experts were cautious to produce the ink-pixel labels as accurately as possible, conservatively avoiding to mark any non-ink pixels. This labeling task took around 4-8 hours for each of the images. We selected 51 fragment images to be labeled for this experiment. The selection was made carefully to accommodate maximum diversities to represent the whole collection (a full list is attached in the Appendix 5).

For the plate image, we used the already labeled fragment images (for that plate) by manually putting them on a transparent layer created on top of the plate image. We used 17 plates to re-train the model (a full list is attached in the Appendix 6).

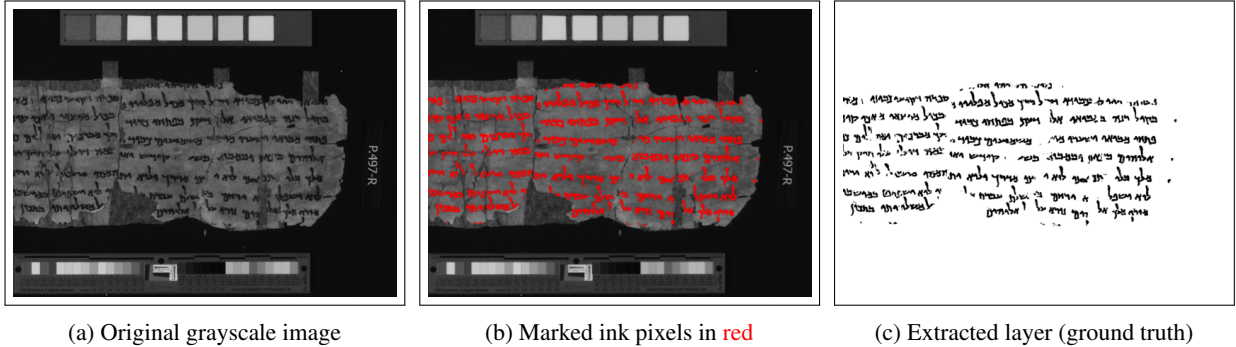


Figure 8: An illustration of the procedure to create ground-truth data from the DSS fragment images using GIMP tool. The image is from fragment 9 of plate 497 (column 2, row 1).

4 Experiments

In this section, we will briefly discuss different aspects of the experiment including the training procedures and the evaluation matrices for quantitative analysis.

4.1 Training

From the labeled fragment images, we use 40 images for training and 11 for testing and evaluation. Each of these images contains one fragment and has been manually binarized. We train this network to minimize the L1-loss. For this, we use Adam optimizer with an initial learning rate of 0.0002. The model is trained for 200 epochs. We train both the proposed framework (BiNet) and the original CGAN model (from pix2pix [19]) on color, grayscale, and fused fragment images separately (from scratch) as described in Subsection 3.4. Additionally, we train our model on the DIBCO datasets from scratch. We used the DIBCO datasets from 2009 to 2014 as training data and the datasets from 2016 and 2017 as test data. For plate images, we used a pre-train model and updated the weights by re-training it for another 200 epochs using the 16 manually labeled plate images. The system runs on a personal workstation with a single GPU (NVIDIA GTX 1060 with 6GB memory; details can be found in the Appendices 8 and 9).

4.2 Evaluation Measures

For the purpose of quantitative analysis, we use evaluation metrics that are commonly used in the (H-)DIBCO [64, 65, 66, 67, 42, 43, 57]. The metrics are suitable in the context of document analysis. We will use four metrics:

F-measure (F), pseudo-F-measure (F_{ps}), peak signal-to-noise ratio ($PSNR$), and distance reciprocal distortion (DRD). Brief descriptions for each of them are provided in the following subsections.

4.2.1 F-measure

F-measure (also F1-score or F-score) is the measure of a test’s accuracy. It is defined as:

$$Fmeasure = \frac{2 \times Recall \times Precesion}{Recall + Precision} \quad (3)$$

where, $Recall = \frac{TP}{TP+FN}$ and $Precision = \frac{TP}{TP+FP}$; TP , FP , FN refer to the True Positive, False Positive, and False Negative values, respectively.

4.2.2 pseudo-F-measure

Pseudo F-measure follows the same formula of F-measure (Eq. 4), but it uses pseudo-Recall and pseudo-precision [71].

$$pseudoFmeasure = \frac{2 \times pseudoRecall \times pseudoPrecesion}{pseudoRecall + pseudoPrecision} \quad (4)$$

Both these pseudo metrics use distance weights with respect to the contour of the ground-truth characters. In the case of pseudo-Recall, the weights of the ground-truth inks are normalized according to the local stroke width. These weights are defined between $[0, 1]$. In the case of pseudo-Precision, the weights are constrained within an area that expands to the background of the ground-truth taking into account the stroke width of the nearest ground-truth component. Inside this area, the weights are greater than one (generally between $[1, 2]$) while outside this area they are equal to one.

4.2.3 Peak Signal-to-Noise Ratio (PSNR)

Peak signal-to-noise ratio is the measure of how close an image is to another one. The higher the value of PSNR, the higher the similarity of the two images that are being compared.

$$PSNR = 10 \log\left(\frac{C^2}{MSE}\right) \quad (5)$$

where, C is the difference between foreground and background, and MSE is the the mean squared error and defined as:

$$MSE = \frac{\sum_{x=1}^M \sum_{y=1}^N (I_{bin}(x, y) - I'_{bin}(x, y))^2}{MN} \quad (6)$$

where, $M \times N$ is the dimension of the image, I_{bin} is the original ground-truth image and I'_{bin} is the test-output of the ground-truth from the model.

4.2.4 Distance Reciprocal Distortion (DRD)

The distance reciprocal distortion metric correlates with the human visual perception system and measures the distortion for all the S -flipped pixels as:

$$DRD = \frac{\sum_{k=1}^S DRD_k}{NUBN} \quad (7)$$

where, DRD_k is the distortion of the k -th flipped pixel that is calculated using 5×5 normalized weight matrix. The weight matrix W_{Nm} is defined by Lu et al. [72], where they used the DRD to measure the visual distortion in binary document images. $NUBN$ is the number of non-uniform 8×8 blocks in the ground-truth image.

5 Results

In this section, we present the experimental results based on four distinct evaluation measures presented in Subsection 4.2. The results are obtained using three different test sets. The **first test set** consists of 11 fragment images of the DSS collection for which the corresponding ground truths were built manually, as described in Subsection 3.6. The selection of the test images was made to accommodate maximum diversity and different degradation (a list is attached in Appendix 5). The **second test set** contains 40 images from H-DIBCO 2016, DIBCO 2017, and H-DIBCO 2018 datasets. Both H-DIBCO 2016 and 2018 have ten handwritten document images each, and DIBCO 2017 has twenty

document images: ten machine-printed and ten handwritten. Finally, the **third test set** consists of 3 RGB-colored full plate images from the DSS collection.

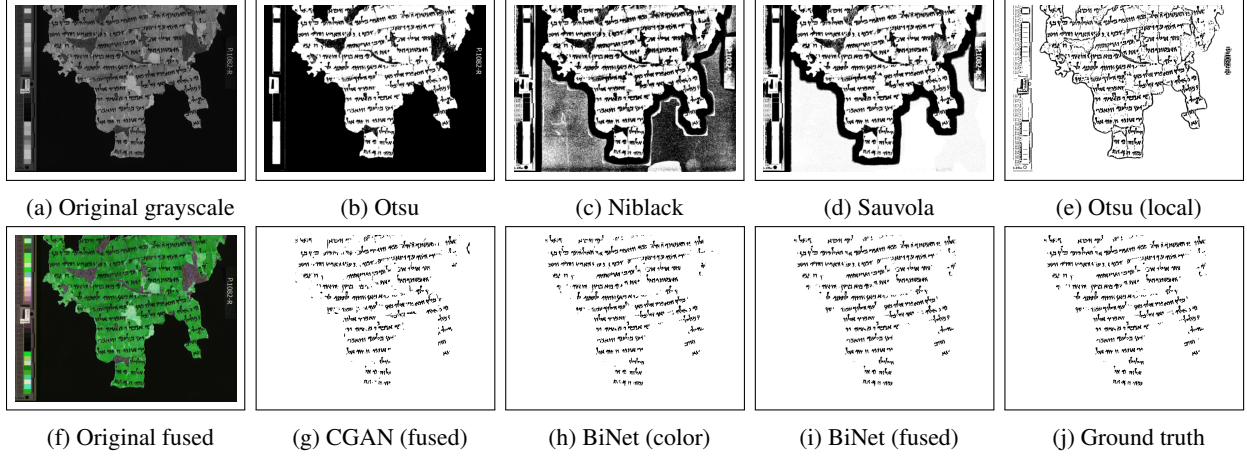


Figure 9: A comparative illustration of test results using different traditional methods along with the proposed BiNet model on fragment 1 of plate 1082. Please note the successful removal of the color-calibrator strip and machine-printed number-tag.

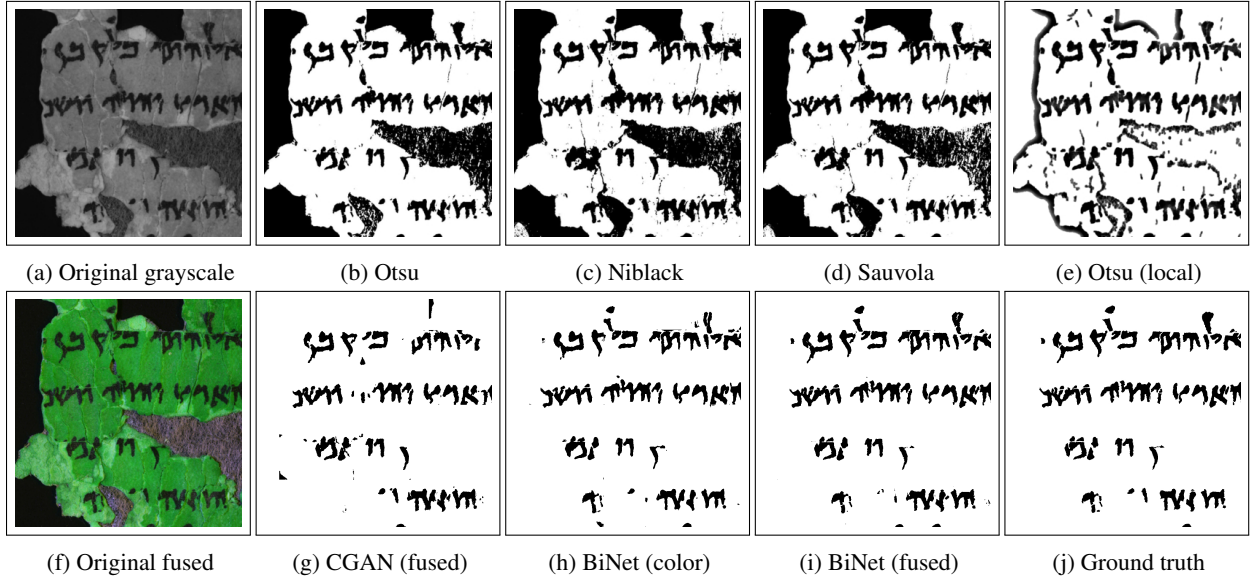


Figure 10: Results of BiNet and traditional methods presented on a zoomed-in part of fragment 1 of plate 1082. The zoomed-in (enlarged) part is taken from the pixel position of (460, 150) with a window size of 780×780 pixels. A visual inspection shows that the BiNet (fused) output is the closest match to the ground truth.

For the **first test-set**, we trained two different models: the proposed BiNet and the conditional adversarial network (CGAN) as proposed in pix2pix (the image-to-image translation work [19]). The reason behind training a CGAN is to evaluate the potential of the adversarial network on learning complex tasks with smaller training data. Additionally, it is worthwhile to present the results of CGAN as the BiNet shares similar generative architecture. Both the CGAN and BiNet model are trained on three different types of images: grayscale, color, and fused (pseudo-color). Thus, we tested the first set of images on six different networks (2 models trained from scratch on three different types of training data).

In order to present a quantitative evaluation, we use four traditional thresholding methods to perform binarization: Otsu [15], Niblack [16], Sauvola [17], and a local implementation of Otsu with 70×70 windows. The quantitative results for the first set, the DSS images, are presented in Table 2. The CGAN models show promising performance compared

to the traditional methods. But all the BiNet models outperform the rest. BiNet on fused is the best performing model in all the four evaluation measures.

Table 2: Detailed evaluation results on the DSS fragment images. The results are presented as $mean \pm std.dev.$ on the whole test set. The proposed BiNet outperforms other methods (best performance in **bold**).

Method	F-measure	pF-measure	PSNR	DRD
Otsu (global) on grayscale	7.3 \pm 3.5	7.3 \pm 3.5	1.6 \pm 0.5	1236 \pm 644.8
Niblack on grayscale	15.7 \pm 8.8	15.9 \pm 9.1	4.9 \pm 1.0	589.4 \pm 317.7
Sauvola on grayscale	19.1 \pm 10.3	19.5 \pm 10.7	6.3 \pm 1.1	429.5 \pm 228.1
Otsu (local) on grayscale	51.3 \pm 13.4	54.9 \pm 13.6	14.8 \pm 1.2	53.7 \pm 25.8
CGAN on grayscale	63.6 \pm 16.6	65.1 \pm 15.7	17.3 \pm 2.2	29.8 \pm 20.2
CGAN on fused	68.4 \pm 18.2	70.5 \pm 16.7	17.8 \pm 2.3	28.6 \pm 23.1
CGAN on color	71.5 \pm 8.7	72.9 \pm 8.4	17.2 \pm 2.0	28.9 \pm 15.2
BiNet on grayscale	80.3 \pm 20.7	82.6 \pm 19.2	20.5 \pm 3.8	18.7 \pm 24.6
BiNet on color	83.5 \pm 9.9	85.8 \pm 9.0	20.3 \pm 2.9	15.2 \pm 14.4
BiNet on fused	86.7 \pm 9.4	89.3 \pm 8.3	21.3 \pm 3.4	13 \pm 14.8

Table 3: Detailed evaluation results on (H-)DIBCO datasets. The BiNet performs equally well with the best-performing methods from each of the competitions (best performance in **bold**). The output images are attached in the Appendices 10, 11, and 12.

	F-measure	pF-measure	PSNR	DRD
H-DIBCO 2016				
Otsu (global)	86.7	90	17.8	5.5
Niblack	56.2	56.3	9.6	57.8
Sauvola	79.9	81.7	14.8	11.4
Best method at H-DIBCO'16 [43]	87.6	91.3	18.1	5.2
BiNet	85.6 (-2.0)	90.7 (-0.6)	18.3 (+0.2)	4.9 (-0.3)
DIBCO 2017				
Otsu (global)	77.7	80.1	13.8	15.8
Niblack	57.4	57.6	8.8	44
Sauvola	80	83	14.3	9.6
Best method at DIBCO'17 [57]	91.0	92.9	18.3	3.4
BiNet	90.9 (-0.1)	93.3 (+0.4)	18.3 (+0.0)	3.6 (+0.2)
H-DIBCO 2018				
Otsu (global)	51.4	53.4	9.7	59.5
Niblack	45.7	45.9	7.7	80.8
Sauvola	56.3	58.7	10.9	36
Best method at H-DIBCO'18 [68]	88.3	90.2	19.1	4.9
BiNet	84.7 (-3.6)	87.1 (-3.1)	17.4 (-1.7)	7.5 (+2.6)

Table 4: Detailed evaluation results on the DSS full plate images. The results are presented as $mean \pm std.dev.$ on the whole test set (best performance in **bold**). The output images are attached in the Appendix 15.

Method	F-measure	pF-measure	PSNR	DRD
Otsu	27.2 \pm 5.8	27.2 \pm 5.8	9.5 \pm 2.1	206.4 \pm 101.8
Niblack	7.3 \pm 5.9	7.3 \pm 5.9	1.9 \pm 1.4	1375.4 \pm 872.7
Sauvola	47.9 \pm 16.1	48 \pm 16.2	13.8 \pm .08	80.9 \pm 43.6
BiNet	86.9 \pm 3.1	91.1 \pm 2.4	22.9 \pm 2.6	6.3 \pm 0.2

For the **second test set** of (H-)DIBCO images, we used the BiNet model only. The quantitative results of the second set are shown in Table 3. For the (H-)DIBCO results, we also present the performances of the winning methods from each year. Though our model is originally designed for DSS images, it shows high performance in the case of (H-)DIBCO images. For the datasets of 2016 and 2017, BiNet outperforms the best-performing method in a couple of evaluation measures. For the cases of 2018-dataset, BiNet is almost as good as the best-performing method (Table 3).

For the **third test set** of RGB-colored full plate images, we used the transfer learning technique on the BiNet model that was initially trained on (H-)DIBCO images. The results are presented in Table 4. The BiNet outperforms the traditional methods by large differences and obtains high evaluation measures. This performance shows the excellent usability of the system on different datasets with a small number of training data by transfer learning techniques.

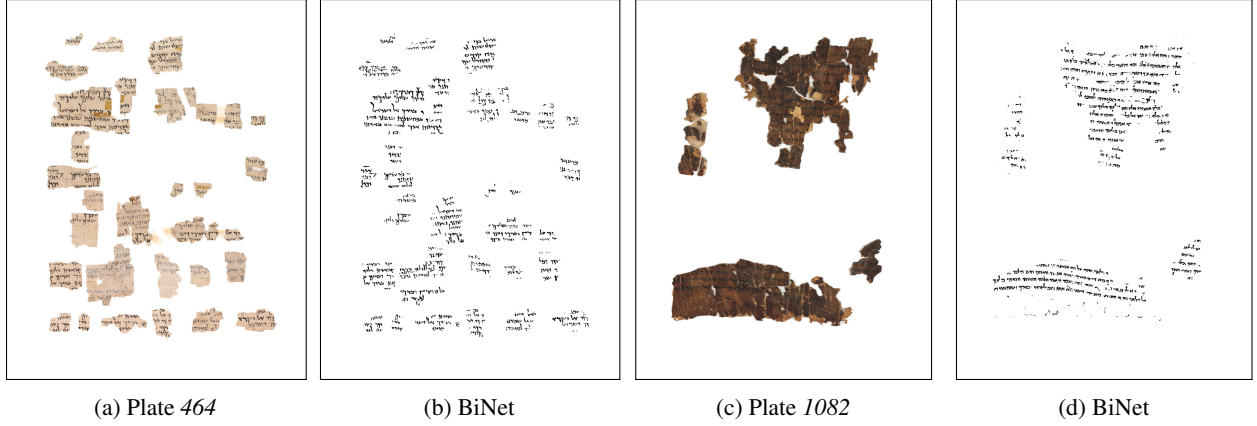


Figure 11: Binarization results of two full plate images from the DSS collection using BiNet (trained on DIBCO images, then updated by transfer learning using sixteen manually labeled plate images).

The resulting images from different methods are presented in Figure 9. For a better qualitative analysis, a zoomed-in portion of the methods can be found in Figure 10. BiNet is extremely good in binarizing the original content and labeling everything else as background. Inside the area of the original content, BiNet is remarkably similar to the ground truth (Figure 10). An interesting finding in the results from the fused images can be seen in Figure 12. During the labeling process of the images, the human expert only labeled the visible inks in the fragments. However, in the fused images, parts of ink might become visible, making it extractable during the binarization process. Thus a fused image binarization can reveal more characters than a visible color image (Figure 12). Please note that the phenomenon might reduce the quantitative performance as we are comparing the output with the ground truth. Nevertheless, this is extraordinary and useful in real applications.

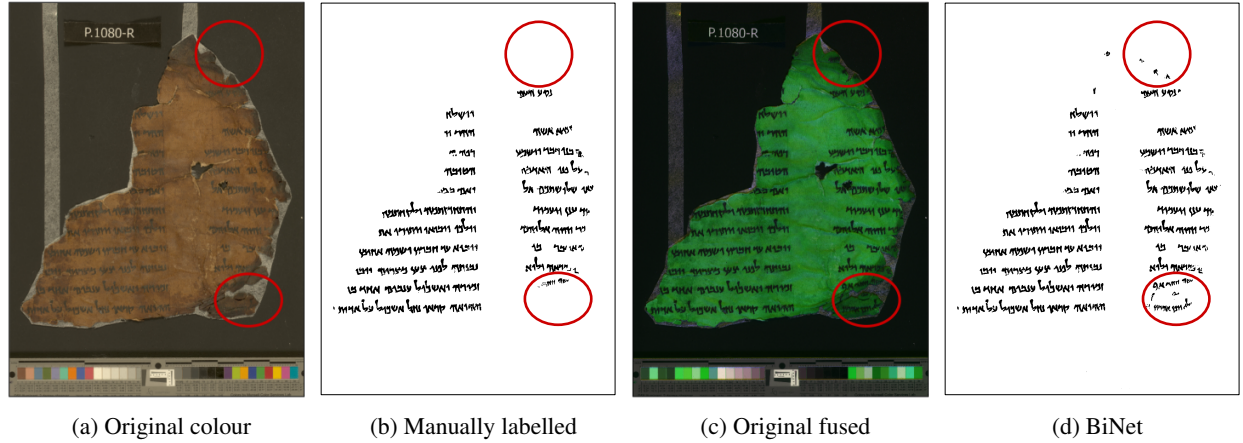


Figure 12: More characters (marked in red circles) are extracted than the ground truth during the binarization process using BiNet on pseudo-color image (fused image) of fragment 6 from plate 1080.

Finally, to test the robustness of the model, we collected some additional materials from different collections to perform the binarization using BiNet. One such image is of the famous Nash Papyrus, which was kindly provided to us by Ben Outhwaite of the Genizah Research Unit at Cambridge University Library. We used the pretrained BiNet model from the DSS color-images to binarize the Nash papyrus and the result is illustrated in Figure 13. The binarization result shows yet another case where the BiNet can extract the original content by segmenting all the irrelevant materials as background pixels. We also performed tests on some grid images produced from several different historical manuscripts (non-DSS) from the Monk system. These results are presented in the Appendix 16.

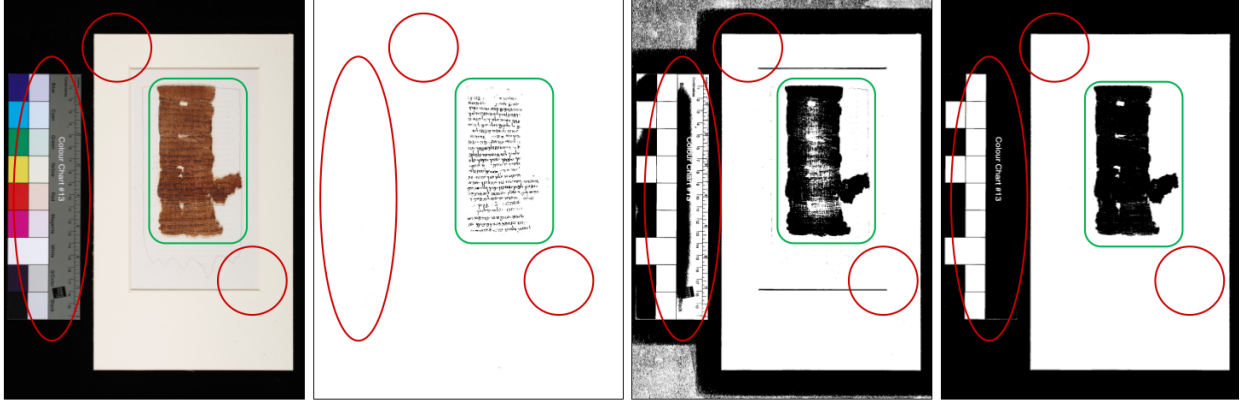


Figure 13: The binarization result from the proposed BiNet on a completely unseen (non-DSS) data (the Nash papyri) to illustrate the robustness of the system. From left to right: the original image, BiNet output, Sauvola, and Otsu. BiNet outputs only the original content from the papyrus fragment (marked in green circle) and removes all the irrelevant components such as the picture frame and the machine-printed color-calibrators (marked in red circles).

6 Conclusions

In this article, we proposed a complete framework, the BiNet, to efficiently binarize one of the most degraded historical manuscript collections, the Dead Sea Scrolls. The method can work with both the full-plate color images and the grayscale intensity images of the individual fragments. The network we used is a variant of U-Net and is capable of learning what is essential and what is irrelevant. Compared to traditional methods, the proposed BiNet can focus on and binarize the original written content of the document with remarkably high performance, which is crucial for getting as close as possible to the original writer to be able to perform writer identification and document dating. For a strictly handwriting based prediction, all non-ink information needs to be removed in these applications. One of the significant features of the network is the ability to segment everything except the original writing-contents into the background semantically. The binarization results of BiNet demonstrate the robustness and multi-purpose usability of the network on different degraded manuscripts of diverse document textures and layouts.

In order to facilitate the training of the neural network, we utilized a simple and effective ground truth labeling technique. We proposed an image fusion technique to produce a pseudo-color image from multi-spectral image-bands. This technique improved the binarization results. In several cases, these improved results from fused images were able to extract more of the original contents than the ground truth itself. Though this phenomenon might lead to lower performance on the quantitative analysis, this additional extraction is much more desirable in the real-world applications. Similar to BiNet, the fusion technique can also be used in different collections if multi-spectral images are available. Though our framework was initially designed for the DSS images, we have performed experiments with (H-)DIBCO datasets as well. The BiNet delivered excellent results for these datasets with high performance-measures, similar to the best-performing methods. This success shows the prospect of using BiNet as a one-off tool for manuscript binarization.

In this article, we have worked on the binarization part only. Additionally, both our ground-truth labeling and binarization results were restrictive. We were careful enough to avoid any automatic filling or smoothing of the characters. This conservative approach made the output of BiNet crisp and accurate to the ink of each character. As most of the historical manuscripts show extreme degradation even on the character level, a reconstruction-based binarization needs to be explored in the future. More research can be done on the image fusion technique itself by fusing more than three channels. In cases of a larger image, we divided the images into smaller ones (256×256). During the binarization process, the border areas sometimes lack in obtaining perfect binarization. This problem can be improved by further work in the area and might be solved by using overlapping patches of images. We used a variant of U-Net due to the low amount of training data with high complexity. If more training data becomes available, a more in-depth network similar to ResNet [73] or even a dense network similar to DenseNet [74] can be explored in the future. Though we propose a complete binarization framework, for now, additional research can always be performed for further improvement of the technique.

7 Acknowledgements

The authors would like to thank Mladen Popović (Principal investigator of the ERC project, Faculty of Theology and Religious Studies, University of Groningen), Eibert Tigchelaar (Research Professor, Faculty of Theology and Religious Studies, KU Leuven) for their valuable inputs and time in labeling and creating the ground truths of the DSS images. The authors would also like to thank Drew Longacre (Post-doctoral fellow, ERC project), Gemma Hayes (Ph.D. candidate, ERC project), Ayhan Aksu (Ph.D. candidate, NWO/FWO project), Marwin Van Dijk (Research assistant, ERC project), and Christine van der Veer (Research assistant, ERC project) for their time and effort in labeling a number of the ground-truth images.

This work has been supported by an ERC Starting Grant of the European Research Council (EU Horizon 2020): *The Hands that Wrote the Bible: Digital Palaeography and Scribal Culture of the DSS* (HandsandBible # 640497).

References

- [1] Jorge Calvo-Zaragoza and Antonio-Javier Gallego. A selectional auto-encoder approach for document image binarization. *Pattern Recognition*, 86:37–47, 2019.
- [2] Marcos Almeida, Rafael Lins, Rodrigo Bernardino, Darlisson Jesus, and Bruno Lima. A new binarization algorithm for historical documents. *Journal of Imaging*, 4(2):27, 2018.
- [3] Georgios Louloudis, Basilios Gatos, Ioannis Pratikakis, and Constantin Halatsis. Text line detection in handwritten documents. *Pattern Recognition*, 41(12):3758–3772, 2008.
- [4] Nabendu Chaki, Soharab Hossain Shaikh, and K Saeed. Exploring image binarization techniques. *sci*, vol. 560, 2014.
- [5] Maruf A Dhali, Sheng He, Mladen Popović, Eibert Tigchelaar, and Lambert Schomaker. A digital palaeographic approach towards writer identification in the dead sea scrolls. In *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods-Volume 1: ICPRAM*, volume 2017, pages 693–702. Scitepress; Setúbal, 2017.
- [6] Sheng He, Marco Wiering, and Lambert Schomaker. Junction detection in handwritten documents and its application to writer identification. *Pattern Recognition*, 48(12):4036–4048, 2015.
- [7] Alaa Sulaiman, Khairuddin Omar, and Mohammad F Nasrudin. Degraded historical document binarization: A review on issues, challenges, techniques, and future directions. *Journal of Imaging*, 5(4):48, 2019.
- [8] Simone Marinai, Emanuele Marino, Francesca Cesarini, and Giovanni Soda. A general system for the retrieval of document images from digital libraries. In *First International Workshop on Document Image Analysis for Libraries, 2004. Proceedings.*, pages 150–173. IEEE, 2004.
- [9] Apostolos Antonacopoulos and Dimosthenis Karatzas. Document image analysis for world war ii personal records. In *First International Workshop on Document Image Analysis for Libraries, 2004. Proceedings.*, pages 336–341. IEEE, 2004.
- [10] Mladen Popović. The ancient library of qumran between urban and rural culture. In *The Dead Sea Scrolls at Qumran and the Concept of a Library*, pages 155–167. BRILL, 2016.
- [11] Alaa Sulaiman, Khairuddin Omar, and Mohammad F Nasrudin. A database for degraded arabic historical manuscripts. In *2017 6th International Conference on Electrical Engineering and Informatics (ICEEI)*, pages 1–6. IEEE, 2017.
- [12] Hubert L Dreyfus and Stuart E Dreyfus. What artificial experts can and cannot do. *AI & society*, 6(1):18–26, 1992.
- [13] Ada Yardeni. A note on a qumran scribe. *New Seals and Inscriptions, Hebrew, Idumean, and Cuneiform, Sheffield*, pages 287–298, 2007.
- [14] Meir Lubetski and Shlomo Moussaieff. *New Seals and Inscriptions, Hebrew, Idumean, and Cuneiform*. Sheffield Phoenix Press, 2007.
- [15] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.
- [16] Wayne Niblack. *An introduction to digital image processing*. Strandberg Publishing Company, 1985.
- [17] Jaakko Sauvola and Matti Pietikäinen. Adaptive document image binarization. *Pattern recognition*, 33(2):225–236, 2000.

- [18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [19] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.
- [20] Josef Kittler and John Illingworth. On threshold selection using clustering criteria. *IEEE transactions on systems, man, and cybernetics*, (5):652–655, 1985.
- [21] Bolan Su, Shijian Lu, and Chew Lim Tan. Binarization of historical document images using the local maximum and minimum. In *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, pages 159–166. ACM, 2010.
- [22] Reza Farrahi Moghaddam and Mohamed Cheriet. Adotsu: An adaptive and parameterless generalization of otsu’s method for document image binarization. *Pattern Recognition*, 45(6):2419–2431, 2012.
- [23] Zhixin Shi, Srirangaraj Setlur, and Venu Govindaraju. Image enhancement for degraded binary document images. In *2011 International Conference on Document Analysis and Recognition*, pages 895–899. IEEE, 2011.
- [24] Norbert Wiener. *Back Matter*. MIT press, 1949.
- [25] Basilios Gatos, Ioannis Pratikakis, and Stavros J Perantonis. Adaptive degraded document image binarization. *Pattern recognition*, 39(3):317–327, 2006.
- [26] Garret D Vo and Chiwoo Park. Robust regression for image binarization under heavy noise and nonuniform background. *Pattern Recognition*, 81:224–239, 2018.
- [27] Hossein Ziaei Nafchi, Reza Farrahi Moghaddam, and Mohamed Cheriet. Phase-based binarization of ancient document images: Model and applications. *IEEE transactions on image processing*, 23(7):2916–2930, 2014.
- [28] Reza Farrahi Moghaddam and Mohamed Cheriet. Beyond pixels and regions: A non-local patch means (nlpm) method for content-level restoration, enhancement, and reconstruction of degraded document images. *Pattern Recognition*, 44(2):363–374, 2011.
- [29] Konstantinos Zagoris and Ioannis Pratikakis. Text detection in natural images using bio-inspired models. In *2013 12th International Conference on Document Analysis and Recognition*, pages 1370–1374. IEEE, 2013.
- [30] Konstantinos Zagoris and Ioannis Pratikakis. Bio-inspired modeling for the enhancement of historical handwritten documents. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 1, pages 287–292. IEEE, 2017.
- [31] Di Lu, Xin Huang, and LiXue Sui. Binarization of degraded document images based on contrast enhancement. *International Journal on Document Analysis and Recognition (IJDAR)*, 21(1-2):123–135, 2018.
- [32] John Canny. A computational approach to edge detection. In *Readings in computer vision*, pages 184–203. Elsevier, 1987.
- [33] Qiang Chen, Quan-sen Sun, Pheng Ann Heng, and De-shen Xia. A double-threshold image binarization method based on edge detector. *Pattern recognition*, 41(4):1254–1267, 2008.
- [34] Marte A Ramírez-Ortegón, Ernesto Tapia, Lilia L Ramírez-Ramírez, Raúl Rojas, and Erik Cuevas. Transition pixel: A concept for binarization based on edge detection and gray-intensity histograms. *Pattern Recognition*, 43(4):1233–1243, 2010.
- [35] Fuxi Jia, Cunzhao Shi, Kun He, Chunheng Wang, and Baihua Xiao. Degraded document image binarization using structural symmetry of strokes. *Pattern Recognition*, 74:225–240, 2018.
- [36] Nicholas R Howe. Document binarization with automatic parameter tuning. *International Journal on Document Analysis and Recognition (IJDAR)*, 16(3):247–258, 2013.
- [37] Yanming Guo, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, and Michael S Lew. Deep learning for visual understanding: A review. *Neurocomputing*, 187:27–48, 2016.
- [38] Chris Tensmeyer and Tony Martinez. Document image binarization with fully convolutional neural networks. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 1, pages 99–104. IEEE, 2017.
- [39] Xujun Peng, Huaigu Cao, and Prem Natarajan. Using convolutional encoder-decoder for document image binarization. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 1, pages 708–713. IEEE, 2017.

- [40] Guillaume Lazzara, Roland Levillain, Thierry Géraud, Yann Jacquélet, Julien Marquegnies, and Arthur Crépin-Leblond. The scribo module of the olena platform: a free software framework for document image analysis. In *2011 International Conference on Document Analysis and Recognition*, pages 252–258. IEEE, 2011.
- [41] Guillaume Lazzara and Thierry Géraud. Efficient multiscale sauvola’s binarization. *International Journal on Document Analysis and Recognition (IJDAR)*, 17(2):105–123, 2014.
- [42] Konstantinos Ntirogiannis, Basilis Gatos, and Ioannis Pratikakis. Icfhr2014 competition on handwritten document image binarization (h-dibco 2014). In *2014 14th International conference on frontiers in handwriting recognition*, pages 809–813. IEEE, 2014.
- [43] Ioannis Pratikakis, Konstantinos Zagoris, George Barlas, and Basilis Gatos. Icfhr2016 handwritten document image binarization contest (h-dibco 2016). In *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 619–623. IEEE, 2016.
- [44] Jean-Christophe Burie, Mickaël Coustaty, Setiawan Hadi, Made Windu Antara Kesiman, Jean-Marc Ogier, Erick Paulus, Kimheng Sok, I Made Gede Sunarya, and Dona Valy. Icfhr2016 competition on the analysis of handwritten text in images of balinese palm leaf manuscripts. In *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 596–601. IEEE, 2016.
- [45] Seyed Morteza Ayatollahi and Hossein Ziaei Nafchi. Persian heritage image binarization competition (phibc 2012). In *2013 First Iranian Conference on Pattern Recognition and Image Analysis (PRIA)*, pages 1–4. IEEE, 2013.
- [46] Jorge Calvo-Zaragoza, Gabriel Vigliensoni, and Ichiro Fujinaga. Pixel-wise binarization of musical documents with convolutional neural networks. In *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, pages 362–365. IEEE, 2017.
- [47] Muhammad Zeshan Afzal, Joan Pastor-Pellicer, Faisal Shafait, Thomas M Breuel, Andreas Dengel, and Marcus Liwicki. Document image binarization using lstm: A sequence learning approach. In *Proceedings of the 3rd International Workshop on Historical Document Imaging and Processing*, pages 79–84. ACM, 2015.
- [48] Florian Westphal, Niklas Lavesson, and Håkan Grahn. Document image binarization using recurrent neural networks. In *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*, pages 263–268. IEEE, 2018.
- [49] Nal Kalchbrenner, Ivo Danihelka, and Alex Graves. Grid long short-term memory. *arXiv preprint arXiv:1507.01526*, 2015.
- [50] Quang Nhat Vo, Soo Hyung Kim, Hyung Jeong Yang, and Gueesang Lee. Binarization of degraded document images based on hierarchical deep supervised network. *Pattern Recognition*, 74:568–586, 2018.
- [51] Sheng He and Lambert Schomaker. Deepotsu: Document enhancement and binarization using iterative deep learning. *Pattern Recognition*, 91:379–390, 2019.
- [52] Fanbo Deng, Zheng Wu, Zheng Lu, and Michael S Brown. Binarizationshop: a user-assisted software suite for converting old documents to black-and-white. In *Proceedings of the 10th annual joint conference on Digital libraries*, pages 255–258. ACM, 2010.
- [53] Hossein Ziaei Nafchi, Seyed Morteza Ayatollahi, Reza Farrahi Moghaddam, and Mohamed Cheriet. An efficient ground truthing tool for binarization of historical manuscripts. In *2013 12th International Conference on Document Analysis and Recognition*, pages 807–811. IEEE, 2013.
- [54] Rachid Hedjam, Hossein Ziaei Nafchi, Reza Farrahi Moghaddam, Margaret Kalacska, and Mohamed Cheriet. Icdar 2015 contest on multispectral text extraction (ms-tex 2015). In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 1181–1185. IEEE, 2015.
- [55] George R Cross and Anil K Jain. Markov random field texture models. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (1):25–39, 1983.
- [56] Bowornrat Sriraman and Lambert Schomaker. Explicit foreground and background modeling in the classification of text blocks in scene images. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 755–759. IEEE, 2015.
- [57] Ioannis Pratikakis, Konstantinos Zagoris, George Barlas, and Basilis Gatos. Icdar2017 competition on document image binarization (dibco 2017). In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 1, pages 1395–1403. IEEE, 2017.
- [58] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

- [59] Daniel Michelsanti and Zheng-Hua Tan. Conditional generative adversarial networks for speech enhancement and noise-robust speaker verification. *arXiv preprint arXiv:1709.01703*, 2017.
- [60] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3, 2013.
- [61] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*, 2014.
- [62] The leon levy dead sea scrolls digital library. <https://www.deadseascrolls.org.il/>. Accessed: 2019-10-16.
- [63] Pnina Shor. The leon levy dead sea scrolls digital library. the digitization project of the dead sea scrolls. In *Digital Humanities in Biblical, Early Jewish and Early Christian Studies*, pages 9–20. BRILL, 2014.
- [64] Basilis Gatos, Konstantinos Ntirogiannis, and Ioannis Pratikakis. Icdar 2009 document image binarization contest (dibco 2009). In *2009 10th International conference on document analysis and recognition*, pages 1375–1382. IEEE, 2009.
- [65] Ioannis Pratikakis, Basilis Gatos, and Konstantinos Ntirogiannis. H-dibco 2010-handwritten document image binarization competition. In *2010 12th International Conference on Frontiers in Handwriting Recognition*, pages 727–732. IEEE, 2010.
- [66] Ioannis Pratikakis, Basilis Gatos, and Konstantinos Ntirogiannis. Icfhr 2012 competition on handwritten document image binarization (h-dibco 2012). In *2012 international conference on frontiers in handwriting recognition*, pages 817–822. IEEE, 2012.
- [67] Ioannis Pratikakis, Basilis Gatos, and Konstantinos Ntirogiannis. Icdar 2013 document image binarization contest (dibco 2013). In *2013 12th International Conference on Document Analysis and Recognition*, pages 1471–1476. IEEE, 2013.
- [68] Ioannis Pratikakis, Konstantinos Zagoris, Panagiotis Kaddas, and Basilis Gatos. Icfhr2018 competition on handwritten document image binarization (h-dibco 2018). In *2018 65th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 489–493. IEEE, 2018.
- [69] Monk - search and annotation tools for handwritten manuscripts. <http://monk.hpc.rug.nl/>. Accessed: 2019-10-16.
- [70] Gnu image manipulation program - gimp (version 2.8.6), 2016.
- [71] Konstantinos Ntirogiannis, Basilis Gatos, and Ioannis Pratikakis. Performance evaluation methodology for historical document image binarization. *IEEE Transactions on Image Processing*, 22(2):595–609, 2012.
- [72] Haiping Lu, Alex C Kot, and Yun Q Shi. Distance-reciprocal distortion measure for binary document images. *IEEE Signal Processing Letters*, 11(2):228–231, 2004.
- [73] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [74] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

Appendices

This section contains additional information and results related to the article. Figure 14 illustrates the box-plots for evaluation metrics. The complete list of train and test images can be found in Table 5, Table 6, and Table 7. A brief configuration of the computing system is provided in Table 8. Figure 15 contains the binarization results of three test images (full-plates) of the DSS collection using BiNet with transfer learning. The BiNet is additionally tested on several different manuscript collections which is illustrated using the binarization of grid-images from Monk in Figure 16. Finally, the binarization results of DIBCO/H-DIBCO datasets are presented at the end of this section.

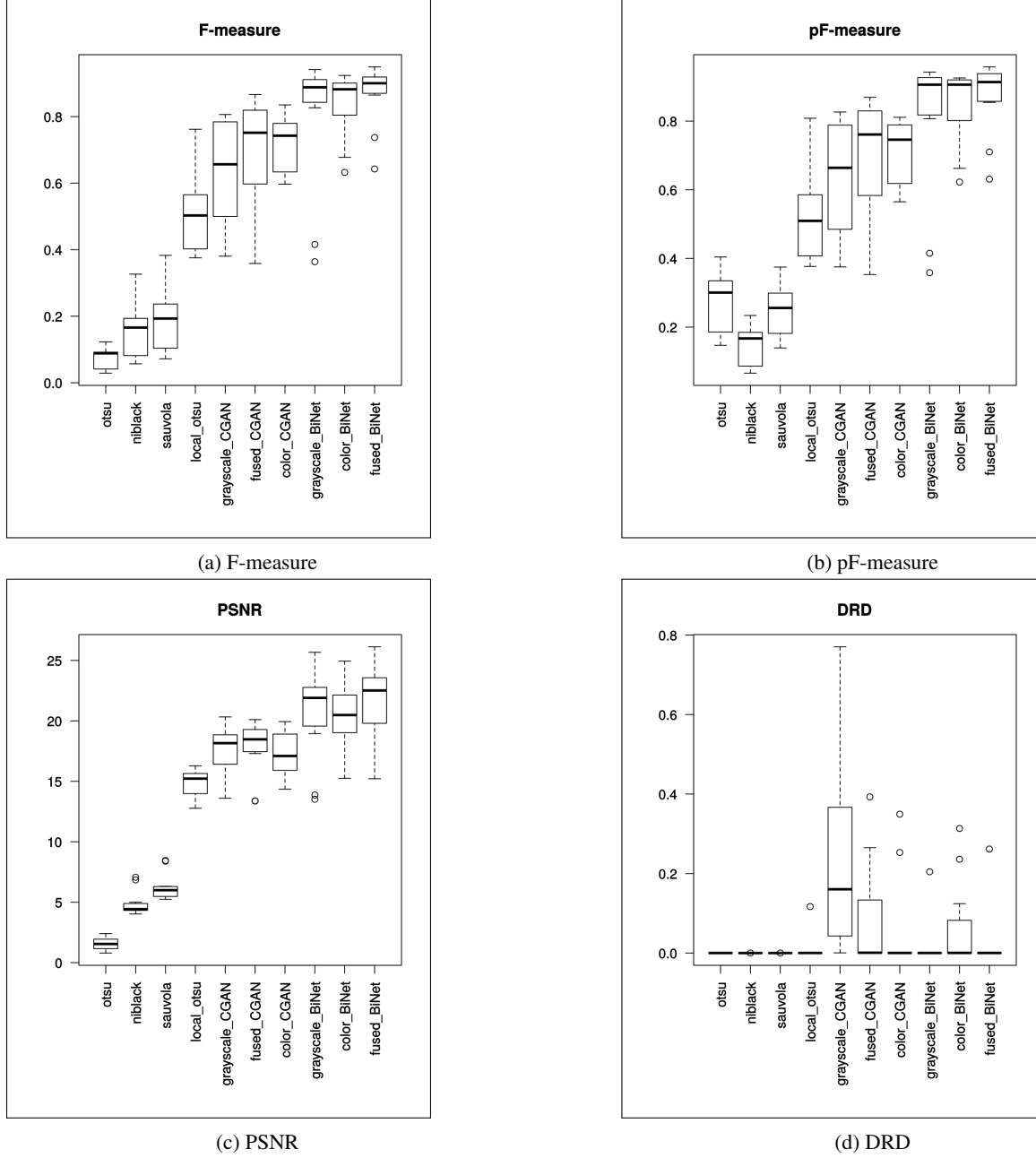


Figure 14: Box-plots showing the distribution of test data (DSS fragment-images) for the four different evaluation metrics.

Table 5: List of train and test images along with material types (for DSS collection of fragment images).

Train-images (columns 1-3)			Test-images
P25-Fg001-R-C01-R01	P497-Fg009-R-C01-R01	P976-Fg001-R-C01-R01	P385-Fg006-R-C01-R01
P25-Fg005-R-C01-R01	P497-Fg009-R-C02-R01	P976-Fg002-R-C01-R01	P385-Fg010-R-C01-R01
P124-Fg004-R-C01-R01	P524-Fg002-R-C01-R01	P976-Fg002-R-C01-R02	P385-Fg011-R-C01-R01
P306-Fg001-R-C01-R01	P580-Fg004-R-C01-R01	P976-Fg003-R-C01-R01	P638-Fg001-R-C02-R01
P307-Fg001-R-C01-R01	P580-Fg004-R-C02-R01	P976-Fg003-R-C01-R02	P638-Fg001-R-C06-R01
P307-Fg003-R-C01-R01	P607-1-Fg001-R-C01-R01	P976-Fg004-R-C01-R01	P834-Fg001-R-C01-R01
P405-Fg001-R-C01-R01	P607-1-Fg001-R-C01-R02	P976-Fg004-R-C01-R02	P834-Fg002-R-C01-R01
P405-Fg001-R-C01-R02	P607-1-Fg001-R-C02-R01	P1001-Fg001-R-C01-R01	P980-Fg001-R-C01-R01
P409-Fg001-R-C01-R01	P607-1-Fg001-R-C02-R02	P1001-Fg002-R-C01-R01	P1080-Fg001-R-C01-R01
P470-Fg001-R-C01-R01	P704-1-Fg002-R-C01-R01	P1001-Fg008-R-C01-R01	P1080-Fg006-R-C01-R01
P470-Fg002-R-C01-R01	P704-1-Fg002-R-C02-R01	P1001-Fg010-R-C01-R01	P1082-Fg001-R-C01-R02
P480-Fg006-R-C01-R01	P807-Fg019-R-C01-R01	P1081A-Fg002-R-C01-R01	
P480-Fg007-R-C01-R01	P807-Fg019-R-C02-R01		
P480-Fg007-R-C02-R01	P807-Fg019-R-C03-R01		
Fragment material types			
	Train-images	Test-images	Total
Parchment	38	9	47
Papyrus	2	2	4
Total	40	11	51

Table 6: List of train (for transfer-learning) and test images (for DSS collection of full-plate images).

Train-images (columns 1-6)					Test-images	
25	307	470	580	807	1081A	1
124	405	480	607-1	976		386
306	409	524	704-1	1001		997
Total number of images:					Train: 16, Test: 3	

Table 7: List of train and test images (from DIBCO/H-DIBCO [64, 65, 66, 67, 42, 43, 57, 68]).

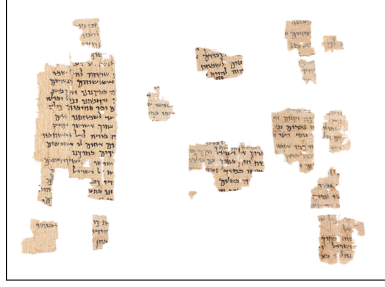
Train-images (columns 1-2)		Test-images
DIBCO 2009	DIBCO 2012	H-DIBCO 2016
DIBCO 2010	DIBCO 2013	DIBCO 2017
DIBCO 2011	H-DIBCO 2014	H-DIBCO 2018
Total number of images:		Train: 76, Test: 40

Table 8: Brief configuration of the work-station.

CPU	Memory	Display card
Intel(R) Core(TM) i5-4590 @ 3.30GHz	description: System memory	GP106 [GeForce GTX 1060 6GB]
size: 3583MHz	size: 7730MiB	vendor: NVIDIA Corporation
capacity: 3700MHz; width: 64 bits		width: 64 bits; clock: 33MHz

Table 9: Time needed to binarize one of the test images (fragment 1, plate 1082)

Grayscale image	RGB-color image	Pseudo-color image
0 min 28.905 sec	0 min 34.991 sec	1 min 5.987 sec



(a) Plate 1



(b) Plate 386



(c) Plate 997



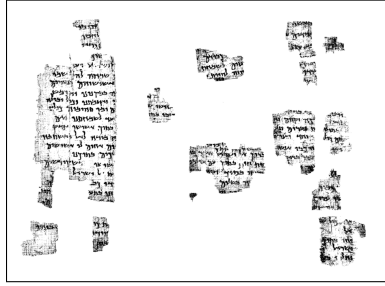
(d) Otsu of Plate 1



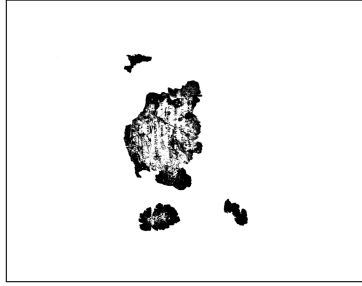
(e) Otsu of Plate 386



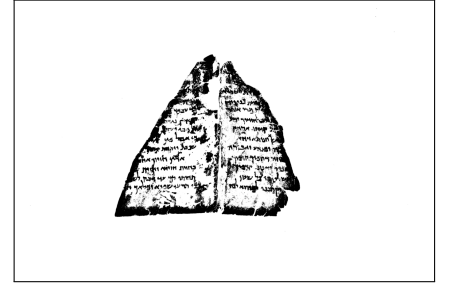
(f) Otsu of Plate 997



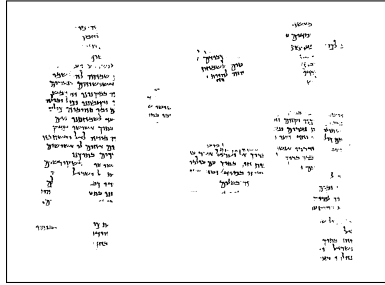
(g) Sauvola of Plate 1



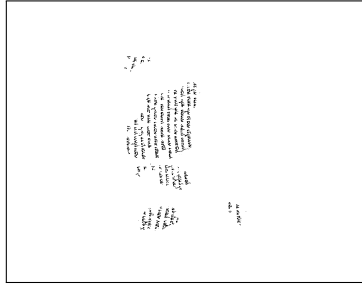
(h) Sauvola of Plate 386



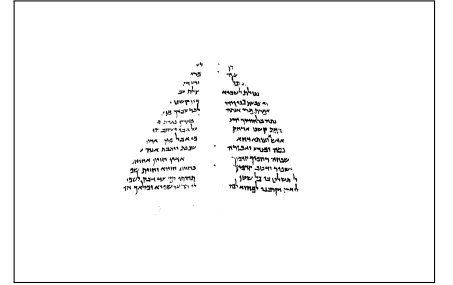
(i) Sauvola of Plate 997



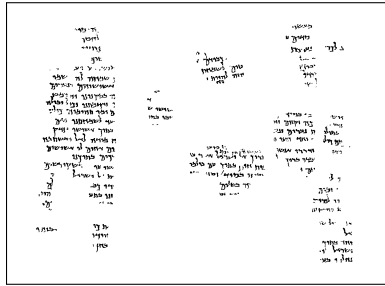
(j) BiNet on Plate 1



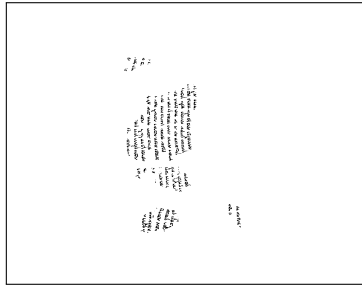
(k) BiNet on Plate 386



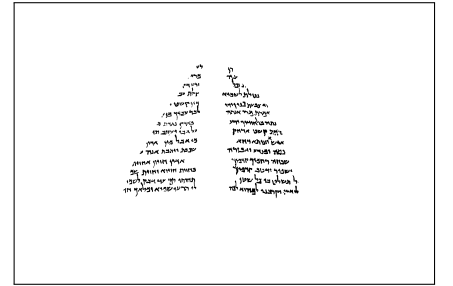
(l) BiNet on Plate 997



(m) Ground-truth of Plate 1



(n) Ground-truth of Plate 386



(o) Ground-truth of Plate 997

Figure 15: Binarization results of three test images (DSS full-plate images) using BiNet (trained on DIBCO images, then updated by transfer learning using sixteen manually labeled plate images).

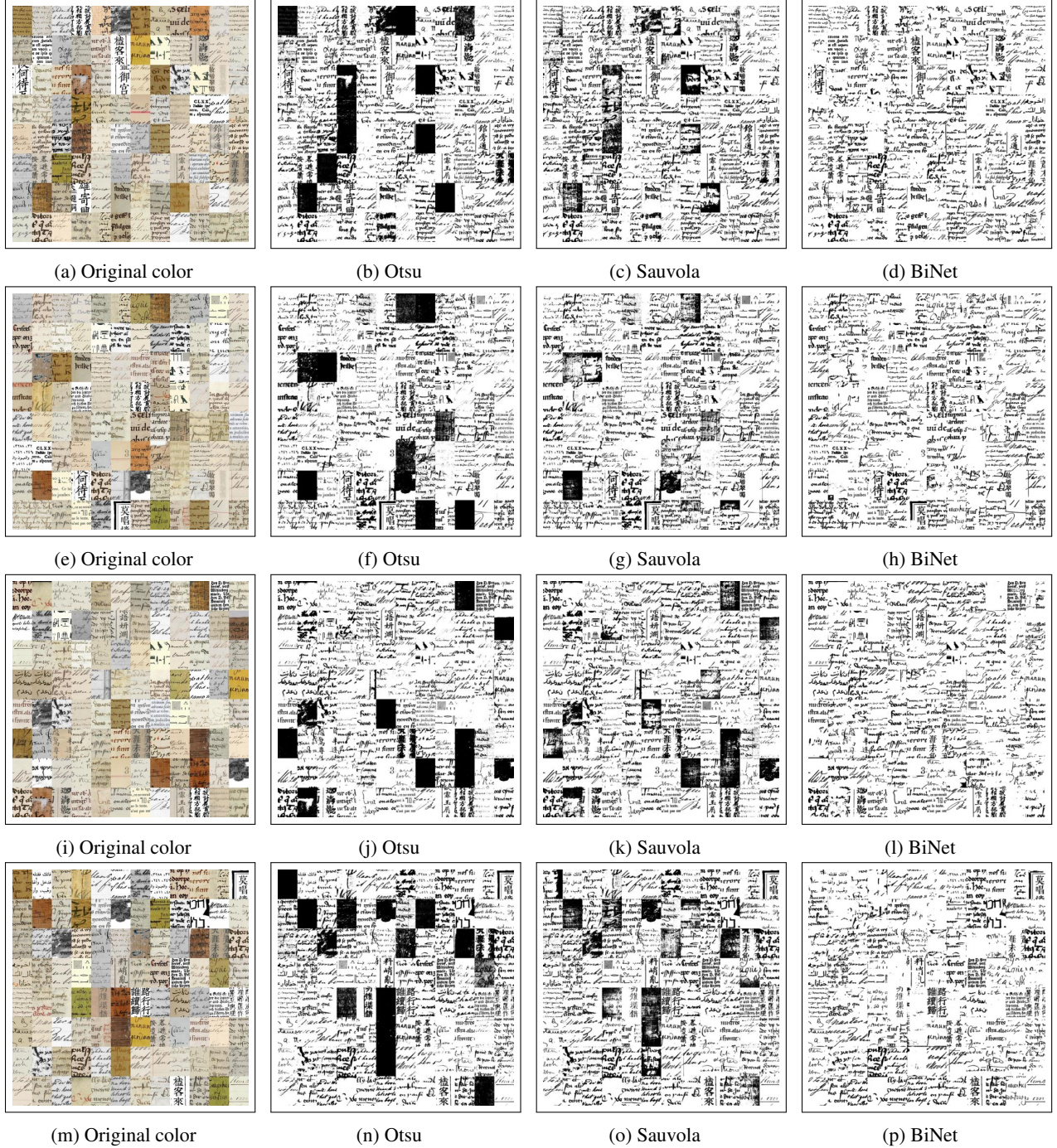


Figure 16: An illustration of BiNet output of four different test images (grid-images) created with various manuscript collections from Monk [69]. The BiNet model used here is trained on DIBCO images. Results from Otsu (global) and Sauvola are presented as well.

		
		
		
		
		
		
		
		
		
		
Original image	BiNet output	Ground-truth

Table 10: Figures of the H-DIBCO 2018 testing dataset along with the binarization results from BiNet.

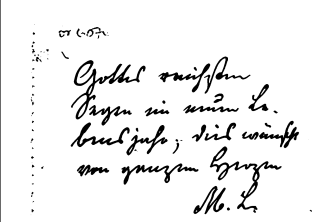
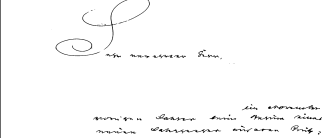
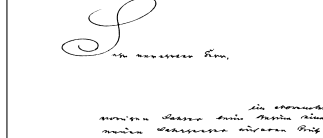
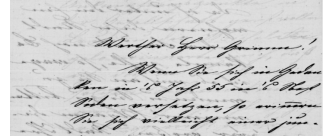
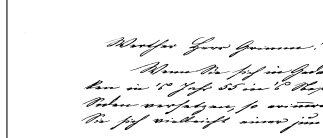
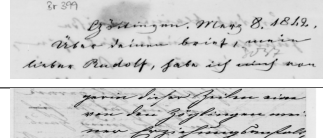
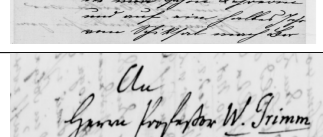
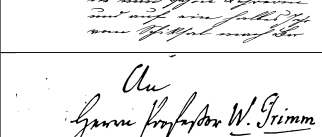
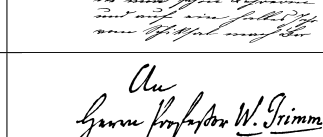
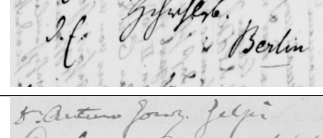
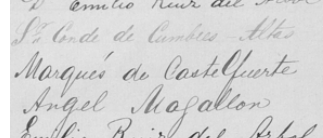
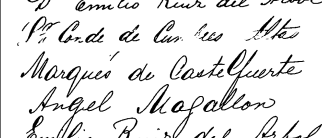
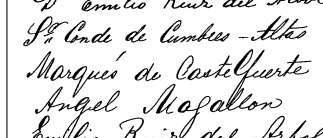

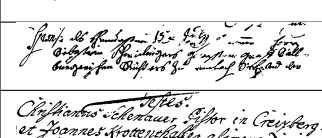
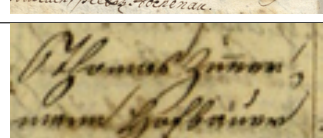
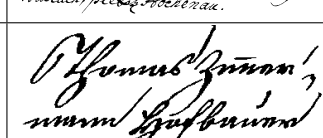
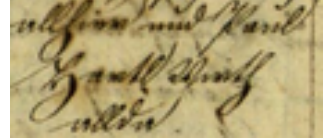
		
		
		
		
		
		
		
		
		
		
Original image	BiNet output	Ground-truth

Table 12: Figures of the H-DIBCO 2016 testing dataset along with the binarization results from BiNet.